INT. J. MATH. EDUC. SCI. TECHNOL., 2004
VOL. 35, NO. 3, 391–402

Enhance-synergism and suppression effects in multiple regression

STAN LIPOVETSKY and W. MICHAEL CONKLIN
GfK Custom Research Inc., 8401 Golden Valley Rd.,
Minneapolis, MN 55427, USA
E-mail: lipovetsky@customresearch.com, mconklin@customresearch.com

(*Received 3 April 2003*)

Relations between pairwise correlations and the coefficient of multiple determination in regression analysis are considered. The conditions for the occurrence of enhance-synergism and suppression effects when multiple determination becomes bigger than the total of squared correlations of the dependent variable with the regressors are discussed. It is shown that such effects can occur just for stochastic relations among the variables that have non-transitive signs of pairwise correlations. Consideration of these problems facilitates better understanding the properties of regression.

1. Introduction

Least squares regression modelling is a regular tool of statistical analysis and prediction for numerous problems in various fields of research and applied estimations. A widely used characteristic of the regression quality is its coefficient of multiple determination. This coefficient corresponds to the share of the variance of a dependent variable y explained by the regression model. It is a convenient measure because its values belong to the interval from zero, for the worst models, to one, for the best. In the case of regression with just one predictor x, the coefficient of multiple determination coincides with the squared correlation between y and x. For a regression by two or more variables $x_1, \ldots, x_n$, the coefficient of multiple determination is a more complicated combination of pair correlations among all xs and y.

We consider properties of the coefficient of multiple determination in its relation to the pair correlations of the dependent variable with regressors. We discuss an interesting possibility of the multiple determination to be bigger than the total of the squared correlations of y with xs. This fascinating feature was described in [1–8]. In social sciences, such effects are discussed in many works [9–12]. Recently this effect and related problems have been discussed further [13–18]. In the statistical literature, this effect is mostly known by the term *enhance-synergism*, so we will use the abbreviation ES for it. In social sciences, the terms *suppression* or *masking* are also used for this effect, and these terms

International Journal of Mathematical Education in Science and Technology
ISSN 0020–739X print/ISSN 1464–5211 online
http://www.tandf.co.uk/journals
DOI: 10.1080/00207390310001658492

correspond to a particular case of the ES effect where some of the predictors are not correlated with the dependent variable, but their participation in the multiple regression increases the quality of the data fitting. We show that the ES effect appears for small correlations, with non-transitive signs of their triplets.

This paper is organized as follows: section 2 describes the case of two predictors and suggests an interpretation of the ES effect. In section 3, we consider the relation between the coefficient of multiple determination and the pairwise correlations for regression with more than two regressors. Section 4 summarizes.

2. Regression with two predictors

Let us briefly describe some relations of the least squares (LS) regression modelling. For centred and normalized by the standard deviations dependent y and n design variables $x_1, \ldots, x_n$ a multiple linear regression model is

$$y = b_1 x_1 + b_2 x_2 + \cdots + b_n x_n + e \tag{1}$$

where e denotes deviations from the model, and b are beta coefficients of the standardized regression. The LS objective of the squared deviations is

$$S^2 = \sum_{i=1}^{N} (y_i - b_1 x_{i1} - b_2 x_{i2} - \cdots - b_n x_{in})^2 \tag{2}$$

where observations are denoted by $i = 1, \ldots, N$. Minimizing (2) by parameters of the model yields a normal system of equations that in matrix form is

$$Cb = r \tag{3}$$

where $C = X'X$ is a matrix of correlations r_{ij} between the predictors, $r = X'y$ is a vector of n correlations r_{yj} of the dependent variable y with each x_j variable, b is the vector of regression coefficients, X denotes design matrix of N by n order, y is the column vector of order N of observations by dependent variable, and X' means the transposed matrix. The solution of this system is

$$b = C^{-1} r \tag{4}$$

where C^{-1} is the inverted correlation matrix. Substituting coefficients of regression (4) into the objective (2) yields the residual sum of squared errors S^2, and the coefficient of multiple determination for multiple regression can be expressed as

$$R^2 = 1 - S^2 = r'b = r'C^{-1} r \tag{5}$$

where r' is a transposed row-vector of correlations of xs with y. The coefficient of multiple determination is always non-negative and less than one.

In the case of one predictor in the model $y = b_1 x_1$, the matrix C in (3) degenerates to a constant, and the solution (4) reduces to

$$b_1 = r_{y1} \tag{6}$$

so the beta coefficient of the pair regression coincides with the correlation of y with the predictor. The coefficient of determination (5) in this case equals the squared correlation:

$$R^2 = r_{y1}^2 \tag{7}$$

Consider the case of two predictors in the regression $y = b_1 x_1 + b_2 x_2$ when the normal system (3) is

$$\begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} r_{y1} \\ r_{y2} \end{pmatrix} \tag{8}$$

so the solution (4) is given by two beta coefficients of regression:

$$b_1 = \frac{r_{y1} - r_{12} r_{y2}}{1 - r_{12}^2} \qquad b_2 = \frac{r_{y2} - r_{12} r_{y1}}{1 - r_{12}^2} \tag{9}$$

The coefficient of multiple determination (5) in this case is

$$R^2 = r_{y1} b_1 + r_{y2} b_2 = \frac{r_{y1}^2 + r_{y2}^2 - 2 r_{y1} r_{y2} r_{12}}{1 - r_{12}^2} \tag{10}$$

If the correlation between x_1 and x_2 is zero, $r_{12} = 0$, and the multiple determination (10) reduces to the equality $R^2 = r_{y1}^2 + r_{y2}^2$. For r_{12} different from zero, intuition suggests the inequality $R^2 \leqslant r_{y1}^2 + r_{y2}^2$, and this is true for most practical data modelling cases. However, the ES effect can occur too, where the multiple determination R^2 is bigger than the sum of the coefficients (7) of determination r_{y1}^2 and r_{y2}^2 in pair regressions $y = b_1 x_1$ and $y = b_2 x_2$, respectively. The inequality

$$R^2 > r_{y1}^2 + r_{y2}^2 \tag{11}$$

shows the ES effect in the explanation of the variance of the dependent variable via two predictors in comparison with simple sum of the explained variance by two separate pair regressions.

The ES effect shows that when working with even simple statistical models, it is better not to rely on intuition but to interpret the derived relations. Horst [9], Kendall and Stuart ([1], Chapter 27.27) and many other authors investigated the possibility of the relation (11). They showed that if the correlation of one cofactor with the criterion variable is zero, for instance, $r_{y2} = 0$, while r_{12} is different from zero, then multiple determination (11) reduces to $R^2 = r_{y1}^2/(1 - r_{12}^2)$ that is always bigger than r_{y1}^2. This effect can be explained by the improved ability to adjust a dependent variable by a linear combination of two predictors with two varying parameters of the regression (1), even if one of the xs is not directly correlated with y. As Weisberg [19] mentioned, the total R^2 can exceed the sum of two squared correlations (11) 'if the two variables interact so that knowing both gives much more information than knowing just one of them. For example, the area of a rectangle may be only poorly determined by either the length or the width alone, but if both length and width are considered in the same model, the area can be determined exactly' (p. 39).

To find the conditions for the occurrence of the ES effect, we construct an index

$$\mathcal{J} = \frac{R^2}{r_{y1}^2 + r_{y2}^2} - 1 = \frac{r_{12}(r_{12} - g)}{1 - r_{12}^2} \tag{12}$$

where g is defined as

$$g = \frac{2 r_{y1} r_{y2}}{r_{y1}^2 + r_{y2}^2} = \frac{G(r_{y1}^2, r_{y2}^2)}{A(r_{y1}^2, r_{y2}^2)}\, sign\,(r_{y1} r_{y2}) \tag{13}$$

By definition (12), the ES effect (11) appears when this index is positive, $\mathcal{J} > 0$. Notation $G(r_{y1}^2, r_{y2}^2)$ and $A(r_{y1}^2, r_{y2}^2)$ in equation (13) correspond to the geometric and arithmetic means of the quantities r_{y1}^2 and r_{y2}^2, so the value of g is below one, and its sign is defined by the sign of the product r_{y1} and r_{y2} in equation (13). The values $\mathcal{J} > 0$ are defined by positive values of the numerator in (12), so solving this quadratic inequality $r_{12}(r_{12} - g) \geqslant 0$ we obtain the following conditions for the ES effect occurrence:

$$\begin{aligned} &\text{if } g \geqslant 0, \quad \text{then } r_{12} \leqslant 0 \text{ or } r_{12} \geqslant g \\ &\text{if } g < 0, \quad \text{then } r_{12} < g \text{ or } r_{12} > 0 \end{aligned} \tag{14}$$

This solution is a generalization of the solution given by Shieh ([13], p. 122).

Three correlations used in (8)–(10) are not independent, they are restricted by the requirement that their total correlation matrix is non-negative definite. So their Gram determinant Δ is non-negative:

$$\Delta = \begin{vmatrix} 1 & r_{12} & r_{y1} \\ r_{12} & 1 & r_{y2} \\ r_{y1} & r_{y2} & 1 \end{vmatrix} = 1 + 2r_{y1}r_{y2}r_{12} - r_{y1}^2 - r_{y2}^2 - r_{12}^2 \geqslant 0 \tag{15}$$

For the given values of any two correlations, the third one will belong to the range satisfying the restriction (15). The quadratic inequality (15) yields the range for r_{12}:

$$r_{y1}r_{y2} - \sqrt{(1 - r_{y1}^2)(1 - r_{y2}^2)} \leqslant r_{12} \leqslant r_{y1}r_{y2} + \sqrt{(1 - r_{y1}^2)(1 - r_{y2}^2)} \tag{16}$$

which is always inside the range from -1 to $+1$. The expression (16) can be obtained from the definition of the partial correlation $r_{12\cdot y}$ between x_1 and x_2 for fixed values of y:

$$r_{12\cdot y} = \frac{r_{12} - r_{y1}r_{y2}}{\sqrt{(1 - r_{y1}^2)(1 - r_{y2}^2)}} \tag{17}$$

Using the borders ± 1 of the interval of possible values on the left-hand side (17), we obtain the inequalities (16) for the range to which a simple correlation r_{12} belongs.

Inequality (11) together with $R^2 \leqslant 1$ means that the restriction $r_{y1}^2 + r_{y2}^2 \leqslant 1$ should hold for the ES cases. Then re-writing the latter inequality as $(r_{y1}r_{y2})^2 \leqslant (1 - r_{y1}^2)(1 - r_{y2}^2)$, taking the square root of this relation, and comparing it with (16), we see that the range of possible r_{12} in the ES cases includes both positive and negative values.

As Mitra [6] observed, for any values of r_{y1} and r_{y2} moving the values of r_{12} towards the bounds in (16), we increase multiple determination R^2 to the maximum possible value 1. Consider the residual sum of squares S^2 (2) that can be presented via the determinant (15) as

$$S^2 = 1 - R^2 = \frac{\Delta}{1 - r_{12}^2} \tag{18}$$

The value $\Delta = 0$ of the determinant (15) defines the margin values in (16) and yields $R^2 = 1$ from (18), which corresponds to the exact linear dependency among three variables. The ratio of multiple determination to the sum $r_{y1}^2 + r_{y2}^2$ equals its maximum value $1/(r_{y1}^2 + r_{y2}^2)$ at the borders of the range (16). The smaller are $|r_{y1}|$ and $|r_{y2}|$, the higher is this ratio.

Let us find some qualitative criteria that help to interpret the occurrence of the ES effect. We can rewrite the relation (10) as

$$R^2 - (r_{y1}^2 + r_{y2}^2) = R^2 r_{12}^2 - 2r_{y1}r_{y2}r_{12} \tag{19}$$

If the product of all three correlations is negative,

$$sign\,(r_{y1}r_{y2}r_{12}) = -1 \tag{20}$$

then the right-hand side of the equality (19) is positive. Then the left-hand side (19) is also positive, so (11) is fulfilled. Thus, (20) is a sufficient condition for the observation of ES effect (11).

The condition (20) means that the values of three coefficients of correlation are of the signs (+, +, −) or (−, −, −). Such sets of signs indicate a specific stochastic relations among three variables and can occur for small (close to zero) correlations. For bigger (further from zero) correlations, we have 'deterministic' sets of signs (+, +, +) and (−, −, +), meaning that if one variable is positively (negatively) correlated with the other, and this latter one is positively (negatively) correlated with the third one, then the first and the third variables will be positively correlated. In [14] (and [16]) it is shown that condition for transitive, 'deterministic' signs corresponds to the inequality $r_{y1}^2 + r_{y2}^2 > 1$. But for this an inequality we cannot have the ES effect (11), because $R^2 \leqslant 1$. So the ES situation (11) can occur only for non-transitive signs of correlations identified by the criterion (20).

Condition (20) corresponds to two larger areas defined by the ranges (14), when $g > 0$ and $-1 < r_{12} < 0$, or when $g < 0$ and $0 < r_{12} < 1$. Another criterion that covers the rest of the range (14), when $g > 0$ and $r_{12} > g$, or when $g < 0$ and $r_{12} < g$, can be obtained as follows. Regrouping in (10) presents the coefficient of multiple determination as

$$R^2 = 1 - \frac{(1 - r_{y1}^2)(1 - r_{y2}^2)}{1 - r_{12}^2} + \frac{(r_{12} - r_{y1}r_{y2})^2}{1 - r_{12}^2} = 1 - (1 - r_{y1}^2)(1 - r_{y2}^2)\frac{1 - r_{12\cdot y}^2}{1 - r_{12}^2} \tag{21}$$

where $r_{12\cdot y}$ is the partial correlation (17). If this partial correlation squared equals the pair correlation r_{12} squared, then (21) is reduced to $R^2 = r_{y1}^2 + r_{y2}^2 - r_{y1}^2 r_{y2}^2$, which is only slightly less than the total $r_{y1}^2 + r_{y2}^2$. So, we see by (21) that if the value of the partial correlation is bigger than that of the simple correlation, or

$$|r_{12\cdot y}| > |r_{12}| \tag{22}$$

then (11) is true and we observe the ES effect. The situation (22) also describes a stochastic effect when a correlation between residuals of x_1 and x_2 obtained in their trends by y (the partial coefficient of correlation $r_{12\cdot y}$) can overcome correlation between these variables x_1 and x_2 themselves. A variable that can make the partial correlation between two other variables larger than their simple correlation is called a masking variable ([1], p. 331 and [3]).

Besides the coefficient of multiple determination, it is interesting to see the ES effect in other regression characteristics. For example, it can be considered for coefficients of regression [2, 10]. Let us consider the behavior of the items of total R^2 known as net effects. The coefficient of multiple determination can be presented as the scalar product of the vectors of beta coefficients and correlations of the dependent variable with all the regressors in the model, and the items of this scalar product are called net effects. Net effects estimate the shares of influence of the regressors on the dependent variable (see more on this characteristic in [20, 21]). For two regressors, the multiple determination is defined in (10), where the items $R_1^2 = r_{y1}b_1$ and $R_2^2 = r_{y2}b_2$ are the net effects of the predictors. Let us look when both net effects are bigger than the corresponding coefficients of determination (7) of the pair regressions. For this reason we construct the ES indices as in (12):

$$\mathcal{J}_1 = \frac{R_1^2}{r_{y1}^2} - 1 = \frac{r_{12}(r_{12} - Q)}{1 - r_{12}^2}, \qquad \mathcal{J}_2 = \frac{R_2^2}{r_{y2}^2} - 1 = \frac{r_{12}(r_{12} - 1/Q)}{1 - r_{12}^2} \tag{23}$$

where the quotient $Q = r_{y2}/r_{y1}$ is defined as in [13]. To solve the system of inequalities $\mathcal{J}_1 > 0$ and $\mathcal{J}_2 > 0$, let us at first take $Q > 0$. Then the first of indices (23) is positive if $r_{12} < 0$ or $r_{12} > Q$, and the second of (23) is positive if $r_{12} < 0$, or $r_{12} > 1/Q$, so the mutual solution is just $r_{12} < 0$ for $Q > 0$. Then consider $Q < 0$: the first of (23) is positive if $r_{12} < Q$, or $r_{12} > 0$, and the second of indices (23) is positive if $r_{12} > 0$ or $r_{12} < 1/Q$, so the mutual solution of both inequalities is $r_{12} > 0$ for $Q < 0$. Then the combined solution for both positive and negative Q can be expressed in one criterion that coincides with (20). The criterion (20) is the necessary and sufficient condition for satisfying the inequalities for both indices (23) to be positive. If each net effect is larger than the correspondent coefficient of determination in the pair regression (7), then total multiple determination satisfies the ES inequality (11). So the criterion (20) is a sufficient condition for the total inequality (11), although it is not necessary for it, and another criterion (22) identifies the additional part of the environment (14) for the ES effect.

We performed numerical runs for a grid of (r_{y1}, r_{y2}, r_{12}) values along the entire range of their possible values corresponding to (16). Both criteria (20) and (22) work very well indicating the cases when the ES effect (11) is observed for multiple determination, and the criterion (20) shows when this effect occurs for the net effects.

3. Regression with many predictors

For a multiple regression with more than two variables, the ES inequality of (11) kind for multiple determination (5) can be presented as

$$r'C^{-1}r > r'r \tag{24}$$

This inequality was studied by Cuadras [22] who used the eigenvalue representation of the correlation matrix to transform (24) to the inequality $\rho'(I - \Lambda)\rho > 0$, where ρ denotes vector of correlations between y and principal components of the regressors, and I and Λ are identity and eigenvalue diagonal matrices, respectively. It is a nice theoretical result, although practically it is easier to compare R^2 directly with the sum of squared pair correlations to see which is bigger.

It can be instructive to consider the following quasi-paradox. Coefficient of multiple determination (5) belongs to 0–1 interval, so always

$$R^2 \leqslant 1 \tag{25}$$

and the better models are of a higher R^2. If the regressors are independent or in a sample they are not correlated, then the correlation matrix C equals identity matrix I, so in (4) $b = r$, or the regression coefficients equal the pair correlations r_{yj} of y with each x_j as in pair model (6). The coefficient of multiple determination (5) in this case reduces to the squared Euclidean norm of the vector of pair correlations,

$$R^2 = r'r = \sum_{j=1}^{n} r_{yj}^2 \tag{26}$$

It seems that for many non-correlated regressors the coefficient (26) can become bigger than one, which contradicts the correct property (25).

To answer this paradox, consider a total matrix of correlations among all xs and y:

$$D = \begin{pmatrix} C & r \\ r' & 1 \end{pmatrix} \tag{27}$$

This square matrix of $(n+1)$th order consists of the main block of the nth order matrix C of correlations among the xs, the border column-vector r and row-vector r' of correlations between y and xs, and a constant 1 in the last cell. Matrix D is a non-negatively definite matrix, because it can be constructed as a cross-product $D = X'X$. Here X denotes a matrix of N by $(n+1)$ order of standardized observations by $n+1$ variables $x_1, \ldots, x_n$, and y.

If correlations among the xs equal zero, then the block C degenerates to the identity matrix I, and D becomes an edge-diagonal matrix:

$$D = \begin{pmatrix} I & r \\ r' & 1 \end{pmatrix} \tag{28}$$

The characteristic equation $|D - \lambda I| = 0$ for the matrix (28) can be obtained in the explicit form using decomposition of the determinant of this matrix by its last row and column, that yields the equation:

$$(1-\lambda)^{n-1}\left((1-\lambda)^2 - \sum_{j=1}^{n} r_{yj}^2\right) = 0 \tag{29}$$

So the eigenvalues λ of the matrix (28) are:

$$\lambda_1 = 1 - \sqrt{\sum\nolimits_{j=1}^{n} r_{yj}^2}, \qquad \lambda_2 = \cdots = \lambda_n = 1, \qquad \lambda_{n+1} = 1 + \sqrt{\sum\nolimits_{j=1}^{n} r_{yj}^2} \tag{30}$$

and among those just the eigenvalue λ_1 can be negative. To ensure λ_1 is non-negative, as is needed for the non-negatively definite matrix (27), the following inequality is fulfilled:

$$\sum_{j=1}^{n} r_{yj}^2 \leqslant 1 \tag{31}$$

Then the property (25) for the coefficient (26) is fulfilled; therefore, there is no paradox with multiple determination in the case of non-correlated xs.

Now we consider a simple estimation of the multiple determination coefficient just using the vector of pair correlations r of y with xs. From the relation (16), let us estimate r_{12} as a mean value of the interval of its possible values, taking $r_{12} = r_{y1}r_{y2}$, when partial correlation (17) equals zero. Such relation corresponds to a Markov triplet of random variables. In this approximation, we substitute all correlations r_{ij} between the predictors via the products of their correlations with the dependent variable. Then we can represent correlation matrix C (3) as follows:

$$C = diag\,(1 - r_{yi}^2) + rr' \tag{32}$$

where *diag* denotes a diagonal matrix consisting of the elements $1 - r_{yi}^2$, and rr' denotes the outer product of the vector r by itself. Matrix C (32) has ones on the diagonal as a regular correlation matrix, and all the non-diagonal elements are $r_{ij} = r_{yi}r_{yj}$.

Applying a well-known Sherman–Morrison formula [23]

$$(A + uv')^{-1} = A^{-1} - \frac{A^{-1}uv'A^{-1}}{1 + u'A^{-1}v} \tag{33}$$

to the matrix (32), we obtain the inverted matrix

$$C^{-1} = diag\,(1 - r_{yi}^2)^{-1} - \frac{diag\,(1 - r_{yi}^2)^{-1} rr'\, diag\,(1 - r_{yi}^2)^{-1}}{1 + r'\, diag\,(1 - r_{yi}^2)^{-1} r} \tag{34}$$

With this matrix (34), the coefficient of multiple determination (5) equals

$$R^2 = r'C^{-1}r = 1 - \frac{1}{1 + r'\, diag\,(1 - r_{yi}^2)^{-1} r} = 1 - \frac{1}{1 + \sum_{i=1}^{n} r_{yi}^2/(1 - r_{yi}^2)} \tag{35}$$

so we can estimate the coefficient of multiple determination having only pair correlations of the dependent variable with the predictors. The approximation (35) has a property (25) of the regular coefficient of multiple determination. Formula (35) can be represented in a more symmetric form using the odds ratio:

$$\frac{R^2}{1 - R^2} = \frac{r_{y1}^2}{1 - r_{y1}^2} + \cdots + \frac{r_{yn}^2}{1 - r_{yn}^2} \tag{36}$$

This shows that a larger number of regressors always increases the value of R^2. If even one $|r_{yj}|$ is closer to one, R^2 also is closer to one. For $n = 1$, multiple determination coincides with the correlation squared, $R^2 = r_{y1}^2$. If all $|r_{yj}| \ll 1$, formula (36) reduces to

$$\frac{R^2}{1 - R^2} = r_{y1}^2 + \cdots + r_{yn}^2 \tag{37}$$

so the multiple determination is bigger than the total of the squared correlations on the right-hand side of (37). We see again that only in the case of 'weak' statistical relations do we observe the ES paradox (11) or (24) for a large number of regressors.

For an explicit example of the ES effect in the multiple regression, let us consider a model (1) with three predictors. The coefficient of multiple determination (5) for $n=3$ can be written explicitly as follows:

$$R^2 = \frac{r_{y1}^2 + r_{y2}^2 + r_{y3}^2 - 2(r_{y1}r_{y2}r_{12} + r_{y1}r_{y3}r_{13} + r_{y2}r_{y3}r_{23})}{\Delta_{123}}$$

$$- \frac{(r_{y1}^2 r_{23}^2 + r_{y2}^2 r_{13}^2 + r_{y3}^2 r_{12}^2) - 2(r_{y1}r_{y2}r_{13}r_{23} + r_{y1}r_{y3}r_{12}r_{23} + r_{y2}r_{y3}r_{12}r_{13})}{\Delta_{123}} \qquad (38)$$

where Δ_{123} is the determinant (15) constructed by correlations among the predictors (so it coincides with (15) up to the change 'y' to '3'). The numerator in the first ratio in (38) consists of the squared correlations and the triplet products, and in the second ratio the numerator contains the quadruplets of the correlations. All pair correlations are less than one by the absolute value, so the input into R^2 of the first ratio is bigger than of the second one. Eliminating the second item, we obtain the main input in the expression (38):

$$R^2 = \frac{r_{y1}^2 + r_{y2}^2 + r_{y3}^2 - 2(r_{y1}r_{y2}r_{12} + r_{y1}r_{y3}r_{13} + r_{y2}r_{y3}r_{23})}{1 + 2r_{12}r_{23}r_{13} - r_{12}^2 - r_{23}^3 - r_{13}^2} \qquad (39)$$

Multiplying equality (39) by its denominator and regrouping the items, we obtain:

$$R^2 - (r_{y1}^2 + r_{y2}^2 + r_{y3}^2) = R^2(r_{12}^2 + r_{23}^2 + r_{13}^2 - 2r_{12}r_{23}r_{13}) \\ - 2(r_{y1}r_{y2}r_{12} + r_{y1}r_{y3}r_{13} + r_{y2}r_{y3}r_{23}) \qquad (40)$$

This expression is similar to the relation (19) extended to three predictors. If all triplet products of correlations at the right-hand side (40) are negative, it guarantees that the right side in total is positive. Then the left-hand side of the relation (40) is positive as well. But it means that the fulfilment of the non-transitivity relation (20) for the correlation triplets in (40) is the sufficient condition of the ES effect:

$$R^2 \geqslant r_{y1}^2 + r_{y2}^2 + r_{y3}^2 \qquad (41)$$

In the particular case that two regressors are not correlated with the dependent variable, $r_{y2} = r_{y3} = 0$, the total expression (38) reduces to

$$R^2 = r_{y1}^2 \left(\frac{1 - r_{23}^2}{1 + 2r_{12}r_{23}r_{13} - r_{12}^2 - r_{23}^2 - r_{13}^2} \right) = \frac{r_{y1}^2}{1 - R_{1.23}^2} \qquad (42)$$

where we use the formula (10) for the coefficient of multiple determination $R_{1.23}^2$ in the regression of the variable x_1 by the regressors x_2 and x_3. By (42), we see that multiple determination in the model (1) by three xs, two of which are not correlated with y, is always bigger than the coefficient of determination in the pair regression of y by the regressor with a non-zero correlation, or $R^2 > r_{y1}^2$. This is the effect of suppression or the masking variables, that is a particular case of the ES effect.

A similar result can be obtained in a general case of n predictors when only one of those is correlated with the dependent variable. If in the least square problem (3) the vector r contains a non-zero first element, $r_{y1} \neq 0$, and all the other $n-1$ elements equal zero, then beta coefficients of regression are defined from (4) as follows:

$$(b_1, b_2, \ldots, b_n)' = r_{y1}\left(C^{11}, C^{12}, \ldots, C^{1n}\right) \qquad (43)$$

where $(C^{11}, C^{12}, \dots, C^{1n})$ are the elements of the first row of the inverted matrix C^{-1} of correlations among the xs. Therefore, the coefficient of multiple determination is defined by only one item in the scalar product (5) of the vector r by vector b (43), that is

$$R^2 = r_{y1}^2 C^{11} = \frac{r_{y1}^2}{1 - R_{1.23\dots n}^2} \tag{44}$$

where we express the diagonal element C^{11} of the inverted matrix via the coefficient of multiple determination $R_{1.23\dots n}^2$ in the regression of the variable x_1 by all the other predictors $x_2, \dots, x_n$. So in this case, the multiple determination (44) is always bigger than the coefficient of determination in the pair regression, or $R^2 > r_{y1}^2$. Again, it is exactly the suppression or masking effect similar to that in (42). It is interesting to note that all coefficients of regression (43) differ from zero, but only the first predictor's net effect is different from zero and equals the total multiple determination.

In a general case, a convenient way to consider the ES effect is as follows. The diagonal elements of an inverted correlation matrix C^{-1} (called variance inflation factors [19]) equal the reciprocal values of the residual sums of squares in the regressions of each variable by all the rest of them. So $(C^{-1})_{jj} = 1/(1 - R_{j.}^2)$, where $R_{j.}^2$ are coefficients of multiple determination in the models of each x_j by all the other xs:

$$x_j = a_{j1.}x_1 + a_{j2.}x_2 + \dots + a_{j,j-1.}x_{j-1} + a_{j,j+1.}x_j + a_{jn.}x_n \tag{45}$$

where $a_{jk.}$ denotes a parameter of jth regression by kth variable among all $n-1$ other xs (using Yule's notation [1]). The non-diagonal elements in any jth row of C^{-1}, taken with the opposite signs and divided by the diagonal element in the same jth row, coincide with the coefficients (45) of the regression of x_j by all the other xs. So, we can represent the inverted correlation matrix (4) as follows:

$$C^{-1} = \begin{pmatrix} \left(1-R_{1.}^2\right)^{-1} & -\left(1-R_{1.}^2\right)^{-1} a_{12.} & -\left(1-R_{1.}^2\right)^{-1} a_{13.} & \cdots & -\left(1-R_{1.}^2\right)^{-1} a_{1n.} \\ -\left(1-R_{2.}^2\right)^{-1} a_{21.} & \left(1-R_{2.}^2\right)^{-1} & -\left(1-R_{2.}^2\right)^{-1} a_{23.} & \cdots & -\left(1-R_{2.}^2\right)^{-1} a_{2n.} \\ \hdashline -\left(1-R_{n.}^2\right)^{-1} a_{n1.} & -\left(1-R_{n.}^2\right)^{-1} a_{n2.} & -\left(1-R_{n.}^2\right)^{-1} a_{n3.} & \cdots & \left(1-R_{n.}^2\right)^{-1} \end{pmatrix} \tag{46}$$

Multiplying (46) by the vector of correlations r between y and xs, we represent coefficients (4) of the regression (1) as

$$b_j = \frac{r_{yj} - (r'_{-j}a_{j.})}{1 - R_{j.}^2} \tag{47}$$

The scalar product in (47) is arranged by the vector $r_{-j} = X'_{-j}y$ of the correlations between y and all the xs except x_j and by the vector $a'_{j.}$ of coefficients in the regression (45). By X'_{-j} we denote the transposed X matrix (used in (3)), but without the jth predictor. The scalar product in numerator (47) is

$$(r'_{-j}a_{j.}) = y'X_{-j}a_{j.} = y'\tilde{x}_j \equiv \tilde{r}_{yj} \tag{48}$$

where $\tilde{x}_j$ denotes the theoretical values of the jth predictor, estimated via its regression (45) by other predictors. So $\tilde{r}_{yj}$ (48) makes sense of the correlation between y and the theoretical value of jth regressor.

Using (47), (48) in (5), we obtain the coefficient of multiple determination for the model (1):

$$R^2 = \sum_{j=1}^{n} \frac{r_{yj}^2}{1 - R_{j.}^2} - \sum_{j=1}^{n} \frac{r_{yj}\tilde{r}_{yj}}{1 - R_{j.}^2} \tag{49}$$

The first sum in the expression (49) is always bigger than $\sum r_{yj}^2$. So, the ES effect is guaranteed if the second sum in (49) is zero or negative. This can occur when the theoretical estimations of pair correlations $\tilde{r}_{yj}$ (48) are of opposite signs to the estimations r_{yj} by empirical data, or exactly in those cases when any variable x_j is weakly related to the other xs, so its prediction by them is poor.

Resuming, the bigger the number n of predictors in the regression (1), the smaller should be their values of correlation with the dependent variable for the occurrence of the ES effect. The natural reason for this is that the condition

$$r_{y1}^2 + r_{y2}^2 + \cdots + r_{yn}^2 < R^2 \leqslant 1 \tag{50}$$

is required for the ES effect. So the average value of the upper bound for each correlation in (50) should be $|r_{yk}| \leqslant 1/\sqrt{n}$, which decreases quickly when n grows. Again, this means that the ES effect appears in situations when the weak predictors in their aggregation increase the quality of fitting of the dependent variable.

4. Summary

We have considered properties of the coefficient of multiple determination in relation to the enhancement-synergism effect. We show that this is a stochastic effect that can be observed in the case of weak statistical relations. Even those predictors poorly related to the dependent variable can play an important role in creating a multiple regression with a high multiple determination, thus providing higher quality analysis, better precision of fitting, and correspondingly better predictive ability. These results (see also [24, 25]) help to a better understanding of the properties of multiple regression.

References

[1] Kendall, M. G., and Stuart, A., 1973, *The Advanced Theory of Statistics*, Vol. 2 (New York: Hafner Publishing).
[2] Currie, I., and Korabinski, A., 1984, *Statistician*, **33,** 283.
[3] Hamilton, D., 1987, *Am. Statistician*, **41,** 129.
[4] Hamilton, D., 1988, *Am. Statistician*, **42,** 90.
[5] Freund, R. J., 1988, *Am. Statistician*, **42,** 89.
[6] Mitra, S., 1988, *Am. Statistician*, **42,** 89.
[7] Bertrand, P. V., and Holder, R. L., 1988, *Statistician*, **37,** 371.
[8] Sharpe, N. R., and Robertson, R. A., 1997, *Am. Statistician*, **51,** 46.
[9] Horst, P., 1941, *The Prediction of Personal Adjustment*, edited by P. Horst (New York: Social Science Research Council), pp. 431–436.
[10] Cohen, J., and Cohen, P., 1975, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (New Jersey: Erlbaum).
[11] Velicer, W. F., 1978, *Educ. Psychol. Meas.*, **38,** 953.
[12] Smith, R. L., Ager, J. W., and Williams, D. L., 1992, *Educ. Psychol. Meas.*, **52,** 17.
[13] Shieh, G., 2001, *Am. Statistician*, **55,** 121.

[14] LANGFORD, E., SCHWERTMAN, N., and OWENS, M., 2001, *Am. Statistician*, **55,** 322.
[15] FRIEDMAN, L., 2002, *Am. Statistician*, **56,** 82.
[16] LIPOVETSKY, S., 2002, *Am. Statistician*, **56,** 341.
[17] KEFENG, XU, and EVERS, P. T., 2003, *Comput. Op. Res.*, **30,** 297.
[18] LYNN, H. S., 2003, *Am. Statistician*, **57,** 58.
[19] WEISBERG, S., 1985, *Applied Linear Regression*, 2nd edn (New York: Wiley).
[20] LIPOVETSKY, S., and CONKLIN, M., 2001, *Comput. Op. Res.*, **28,** 1333.
[21] LIPOVETSKY, S., and CONKLIN, M., 2001, *Appl. Stoch. Models Business Ind.*, **17,** 319.
[22] CUADRAS, C. M., 1993, *Am. Statistician*, **47,** 256.
[23] HARVILLE, D. A., 1997, *Matrix Algebra from a Statistician' Perspective* (New York: Springer).
[24] LIPOVETSKY, S., and CONKLIN, M., 2001, Regression as Weighted Mean of Partial Lines: Interpretation, Properties, and Extensions, *International Journal of Mathematical Education in Science and Technology*, **32,** 697–706.
[25] LIPOVETSKY, S., and CONKLIN, M., 2003, A Model for Considering Multicollinearity, *International Journal of Mathematical Education in Science and Technology*, **34,** 771–777.